Short Course on Multilevel Methods for PDE-Constrained Optimization and Variational Inequalities

TECHNISCHE UNIVERSITÄT DARMSTADT

Stefan Ulbrich
Department of Mathematics
TU Darmstadt

Joint work with Stefanie Bott, J. Carsten Ziems and in parts with Wolfgang Hess, Debora Clever, Jens Lang, TU Darmstadt.

Supported by DFG SPP 1253 and by GSC 233 “Computational Engineering”.

Outline Short Course

TECHNISCHE UNIVERSITÄT DARMSTADT

- Part 1: Multilevel methods for PDE-constrained optimization based on adaptive discretizations and reduced order models
- Part 2: Multilevel methods and reduced order models for the optimization with variational inequalities

Outline Part 1

- Problem setting and basic concept
- Adaptive multilevel methods for time-dependent PDE-constrained optimization with control constraints
- Implementation of the error control criteria by error estimators
- Extension to reduced order models with error estimation
- Numerical results

PD(A)E-constrained Optimization Problem

TECHNISCHE UNIVERSITÄT DARMSTADT

PDE-constrained Optimization Problem (P):

$$\min_{y \in Y, u \in U} \; J(y,u) \quad \text{subject to} \quad C(y,u) = 0, \quad u \in U_{ad}.$$

Y	State space,	$U_{ad} \subset U$	Convex, closed set,
U	Control space (Hilbert space),	$J : Y \times U \to \mathbb{R}$	Objective function,
		$C : Y \times U_{ad} \to \Lambda^*$	State equation.

Working assumptions:

- $C(y,u) = 0$ represents a (system of) PD(A)E('s) with initial and/or boundary conditions.
- For any $u \in U_{ad}$ the state equation $C(y,u) = 0$ has a unique solution $y = y(u) \in Y$.
- J, C are $\mathcal{C}^1$ and $C_y(y,u) \in \mathcal{L}(Y, \Lambda^*)$ has a bounded inverse.

Simple Example: Parabolic Optimal Control Problem with Control Constraints

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min J(y,u) := \frac{1}{2}\|y(T)-y_d\|^2_{L^2(\Omega)} + \frac{1}{2}\|y-y_Q\|^2_{L^2(Q)} + \frac{\alpha}{2}\|u\|^2_{L^2((0,T)\times\partial\Omega)}$$

$$\begin{aligned}
\text{subject to} \quad & y_t - \Delta y = 0 && \text{in } Q := (0,T)\times\Omega,\\
& \partial_n y + y = u && \text{on } (0,T)\times\partial\Omega,\\
& y(0) = y_0 && \text{in } \Omega\\
& a \le u \le b
\end{aligned}$$

$\Omega \subset \mathbb{R}^2,\ T, \alpha > 0,$

$Y = W(0,T) = \{y \in L^2(0,T;H^1(\Omega)) \ : \ y_t \in L^2(0,T;(H^1)^*(\Omega))\},$

$U = L^2((0,T)\times\partial\Omega),$

$y_0 \in L^2(\Omega), \quad y_d \in L^\infty(\Omega), \quad y_Q \in L^\infty(Q),$

$a, b \in L^\infty((0,T)\times\partial\Omega), \quad a < b.$

Multilevel Method for (P)

TECHNISCHE UNIVERSITÄT DARMSTADT

PDE-constrained Optimization Problem (P):

$$\min_{y \in Y, u \in U} \ J(y, u) \quad \text{subject to} \quad C(y, u) = 0, \quad u \in U_{ad}.$$

Adaptive multilevel solver for (P)

- uses a **hierarchy of adaptive (FE) approximations** of (P)
- approximates (FE) discretizations with **reduced order models**
- controls accuracy by **error estimators**
- converges globally and controls approximation errors

Hierarchy of Approximations

TECHNISCHE UNIVERSITÄT DARMSTADT

Infinite-Dimensional Problem: "Truth"

PDE-constrained Optimization Problem (P):

$$\min_{y\in Y, u\in U} J(y,u) \quad \text{s. t.} \quad C(y,u)=0, \quad u\in U_{ad}.$$

Adaptive FE Discretization: "FE-Truth" We generate adaptive approximations

$$(\mathbf{P}_{h_k}) \quad \min_{y^{h_k}\in Y_{h_k}, u^{h_k}\in U_{h_k}} J^{h_k}(y^{h_k},u^{h_k}) \quad \text{s. t.} \quad C^{h_k}(y^{h_k},u^{h_k})=0, \quad u^{h_k}\in U_{ad}^{h_k}.$$

Later: Adaptive Reduced Order Model Approximation:

$$(\mathbf{P}^r_{h_k}) \quad \min_{y^{h_k,r}\in Y^r_{h_k}, u^{h_k,r}\in U^r_{h_k}} J^{h_k,r}(y^{h_k,r},u^{h_k,r}) \text{ s. t. } C^{h_k,r}(y^{h_k,r},u^{h_k,r})=0, u^{h_k,r}\in U_{ad}^{h_k,r}.$$

Challenges

Questions:

- When to proceed to a more accurate discretization?
- How to choose the next adaptive discretization?
- How to invoke reduced order models?

Related Work

- POD-based optimization: Hinze, Kunisch, Sachs, Tröltzsch, Volkwein,...
- Trust-region POD: Arian, Fahl, Sachs, Schu et al., OSPOD: Volkwein et al.
- Inexact function/constraint evaluations: Pironneau, Polak; Heinkenschloss, Ridzal, Vicente; Nocedal; S.U.; Griewank, Walther; Hintermüller; Tröltzsch.
- Recursive multilevel optimization methods: Gratton, Sartenaer, Toint
- A posteriori error estimators: Bank, Bartels, Becker, Carstensen, Debrabant, Lang, Funken, Hoppe, Rannacher, Siebert, Süli, Zienkiewicz, Zhu, ...
- A posteriori error estimators in optimal control:
  - Goal oriented: Becker, Rannacher; Becker, Kapp, Rannacher; Meidner, Vexler; Vexler, Wollner; Deckelnick, Günther, Hinze; Wollner; Hintermüller, Hoppe
  - Residual type: Li, Liu, Ma, Tang; Liu, Yan; Hintermüller, Hoppe, Iliash, Kieweg; Rösch, Siebert; S.U.
- Error control for ROMs: Canuto, Dahmen, Grepl, Gunzburger, Heinkenschloss, Hinze, Iapichino, Maday, Patera, Quarteroni, Rozza, Sorensen, Tröltzsch, Urban, deVore, Volkwein, Willcox,...

Outline

TECHNISCHE UNIVERSITÄT DARMSTADT

- Problem setting and basic concept
- Adaptive multilevel methods for time-dependent PDE-constrained optimization with control constraints
- Implementation of the error control criteria by error estimators
- Extension to reduced order models with error estimation
- Numerical results

Optimization Methods:
Full Space Approach vs. Reduced Approach

TECHNISCHE UNIVERSITÄT DARMSTADT

Full space approach:

Consider (P) as optimization problem in (y, u).

PDE-constrained Optimization Problem (P):

$$\min_{y \in Y, u \in U} J(y,u) \quad \text{s. t.} \quad C(y,u) = 0, \quad u \in U_{ad}.$$

- Solves PDE and optimization problem at the same time
- Linearizes PDE-constraint in every optimization iteration
- Attractive in particular for stationary problems

Optimization Methods:
Full Space Approach vs. Reduced Approach

TECHNISCHE UNIVERSITÄT DARMSTADT

Reduced approach:

Reduce (P) to an optimization problem in u.

Reduced Problem:

$$\min_{u \in U} \quad \widehat{J}(u) := J(y(u), u) \quad \text{s. t.} \quad u \in U_{ad},$$

where $y(u)$ is the unique PDE solution: $C(y(u), u) = 0$.

- Existing PDE solvers usable in a modular way
- Attractive in particular for time dependent problems

In this talk:

We concetrate on multilevel methods for the reduced approach.

Variants for the full space approach are also available (multilevel SQP methods).

Necessary Optimality Conditions

TECHNISCHE UNIVERSITÄT DARMSTADT

PDE-constrained Optimization Problem (P):

$$\min_{y \in Y, u \in U} \quad J(y,u) \quad \text{s. t.} \quad C(y,u) = 0, \quad u \in U_{ad}.$$

Lagrangian Function:

$$L(y,u,\lambda) = J(y,u) + \langle \lambda, C(y,u) \rangle_{\Lambda,\Lambda^*}$$

Necessary Optimality Conditions:

Assume that our assumptions on J, C hold. If $(\bar{y}, \bar{u})$ is local optimal solution of (P) then there exists a Lagrange multiplier (adjoint state) $\bar{\lambda} \in \Lambda$ with

$$\begin{aligned} C(\bar{y},\bar{u}) &= 0 && \text{(State equation)}, \\ L_y(\bar{y},\bar{u},\bar{\lambda}) = J_y(\bar{y},\bar{u}) + C_y^*(\bar{y},\bar{u})\bar{\lambda} &= 0 && \text{(Adjoint equation)} \\ \bar{u} \in U_{ad}, \quad (L_u(\bar{y},\bar{u},\bar{\lambda}), u - \bar{u})_{U^*,U} &\geq 0 \ \ \forall\, u \in U_{ad} && \text{(Stationarity)} \end{aligned}$$

Necessary Optimality Conditions

TECHNISCHE UNIVERSITÄT DARMSTADT

PDE-constrained Optimization Problem (P):

$$\min_{y\in Y, u\in U} \quad J(y,u) \quad \text{s. t.} \quad C(y,u)=0, \quad u\in U_{ad}.$$

Lagrangian Function:

$$L(y,u,\lambda) = J(y,u) + \langle \lambda, C(y,u)\rangle_{\Lambda,\Lambda^*}$$

Necessary Optimality Conditions:

Assume that our assumptions on J, C hold. If $(\bar{y},\bar{u})$ is local optimal solution of (P) then there exists a Lagrange multiplier (adjoint state) $\bar{\lambda}\in\Lambda$ with

$$\begin{aligned} C(\bar{y},\bar{u}) &= 0 && \text{(State equation)},\\ L_y(\bar{y},\bar{u},\bar{\lambda}) = J_y(\bar{y},\bar{u}) + C_y^*(\bar{y},\bar{u})\bar{\lambda} &= 0 && \text{(Adjoint equation)}\\ \bar{u} - P_{U_{ad}}(\bar{u} - \nabla_u L(\bar{y},\bar{u},\bar{\lambda})) &= 0 && \text{(Stationarity)} \end{aligned}$$

Time-dependent PDE-Constrained Optimization

TECHNISCHE UNIVERSITÄT DARMSTADT

For **time-dependent PDEs**

- solving the

  linearized PDE

  or the

  nonlinear PDE

  accurately on the discrete level (e.g. with Rosenbrock methods) has about the **same computational costs**

This motivates:

We consider a multilevel optimization method based on the reduced problem.

Reduced Optimization Problem

TECHNISCHE UNIVERSITÄT DARMSTADT

Reduced Problem:

$$\min_{u \in U} \ \widehat{J}(u) := J(y(u), u) \quad \text{s. t.} \quad u \in U_{ad},$$

where $C(y(u), u) = 0$.
Differentiating the identity

$$\widehat{J}(u) = J(y(u), u) = J(y(u), u) + \langle \lambda, C(y(u), u) \rangle_{\Lambda, \Lambda^*} = L(y(u), u, \lambda)$$

we obtain the well known

Adjoint representation of the reduced derivative:

- Solve $C(y, u) = 0$ for y. $\rightsquigarrow \widehat{J}(u) = J(y, u)$
- Solve $L_y(y, u, \lambda) = 0$ for λ (Adjoint equation). $\rightsquigarrow \widehat{J}'(u) = L_u(y, u, \lambda)$

Hessian: $\widehat{J}''(u)s$ requires additional linearized state and adjoint solve.

Necessary Optimality Conditions Revisited

TECHNISCHE UNIVERSITÄT DARMSTADT

PDE-constrained Optimization Problem (P):

$$\min_{y \in Y, u \in U} \quad J(y,u) \quad \text{s. t.} \quad C(y,u) = 0, \quad u \in U_{ad}.$$

Lagrangian Function:

$$L(y,u,\lambda) = J(y,u) + \langle \lambda, C(y,u) \rangle_{\Lambda,\Lambda^*}$$

Necessary Optimality Conditions:

Assume that our assumptions on J, C hold. If $(\bar{y}, \bar{u})$ is local optimal solution of (P) then there exists a Lagrange multiplier (adjoint state) $\bar{\lambda} \in \Lambda$ with

$$\begin{aligned} C(\bar{y},\bar{u}) &= 0 && \text{(State equation)}, \\ L_y(\bar{y},\bar{u},\bar{\lambda}) = J_y(\bar{y},\bar{u}) + C_y^*(\bar{y},\bar{u})\bar{\lambda} &= 0 && \text{(Adjoint equation)} \\ \bar{u} - P_{U_{ad}}(\bar{u} - \nabla_u L(\bar{y},\bar{u},\bar{\lambda})) &= 0 && \text{(Stationarity)} \end{aligned}$$

Necessary Optimality Conditions Revisited

TECHNISCHE UNIVERSITÄT DARMSTADT

PDE-constrained Optimization Problem (P):

$$\min_{y \in Y, u \in U} \quad J(y,u) \quad \text{s. t.} \quad C(y,u) = 0, \quad u \in U_{ad}.$$

Lagrangian Function:

$$L(y,u,\lambda) = J(y,u) + \langle \lambda, C(y,u) \rangle_{\Lambda,\Lambda^*}$$

Necessary Optimality Conditions:

Assume that our assumptions on J, C hold. If $(\bar{y}, \bar{u})$ is local optimal solution of (P) then there exists a Lagrange multiplier (adjoint state) $\bar{\lambda} \in \Lambda$ with

$$\begin{aligned} C(\bar{y},\bar{u}) &= 0 && \text{(State equation)},\\ L_y(\bar{y},\bar{u},\bar{\lambda}) = J_y(\bar{y},\bar{u}) + C_y^*(\bar{y},\bar{u})\bar{\lambda} &= 0 && \text{(Adjoint equation)}\\ \bar{u} - P_{U_{ad}}(\bar{u} - \nabla \widehat{J}(\bar{u})) &= 0 && \text{(Stationarity)} \end{aligned}$$

Example: Optimal Control of 2D Navier-Stokes Equations

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min J(y,u) := \tfrac{1}{2}\|y - y_Q\|^2_{L^2(Q)} + \tfrac{\alpha}{2}\|u\|^2_{L^2((0,T)\times\Omega_c)}$$

$$\text{subject to} \quad \begin{aligned} y_t + y\cdot\nabla y - \nu\Delta y + \nabla p &= 1_{\Omega_c} u && \text{in } Q := (0,T)\times\Omega,\\ \operatorname{div} y &= 0 \\ y &= 0 && \text{on } (0,T)\times\partial\Omega,\\ y(0) &= y_0 && \text{in } \Omega \\ a \le u &\le b \end{aligned}$$

$\Omega_c \subset \Omega \subset \mathbb{R}^2,\ T, \alpha > 0,$

$V = \overline{\{v \in C_0^\infty(\Omega) : \operatorname{div} v = 0\}}^{H_0^1}$, $H = \overline{\{v \in C_0^\infty(\Omega) : \operatorname{div} v = 0\}}^{L^2}$,

$Y = W(0,T) = \{y \in L^2(0,T;V) : y_t \in L^2(0,T;V^*)\}$,

$U = L^2((0,T)\times\Omega_c)$,

$y_0 \in H, \quad y_Q \in W(0,T)$,

$a, b \in L^\infty((0,T)\times\Omega_c), \quad a < b.$

Example: Optimal Control of 2D Navier Stokes Equations

TECHNISCHE UNIVERSITÄT DARMSTADT

Objective function: $J(y,u) := \frac{1}{2}\|y - y_Q\|^2_{L^2(Q)} + \frac{\alpha}{2}\|u\|^2_{L^2((0,T)\times\Omega_c)}$

State equation: $y \in Y = W(0,T), \quad C(y,u) = 0$ is given by

$$\langle y_t + y\cdot\nabla y - 1_{\Omega_c} u, v\rangle_{L^2(0,T;V^*),L^2(0,T;V)} + \nu(\nabla y, \nabla v)_{L^2} = 0 \quad \forall\, v \in L^2(0,T;V),$$
$$(y(0) - y_0, v_0)_{L^2} = 0 \quad \forall\, v_0 \in H.$$

Example: Optimal Control of 2D Navier Stokes Equations

TECHNISCHE UNIVERSITÄT DARMSTADT

Objective function: $J(y,u) := \frac{1}{2}\|y - y_Q\|^2_{L^2(Q)} + \frac{\alpha}{2}\|u\|^2_{L^2((0,T)\times\Omega_c)}$

State equation: $y \in Y = W(0,T), \quad C(y,u) = 0$ is given by

$$\langle y_t + y \cdot \nabla y - 1_{\Omega_c} u, v \rangle_{L^2(0,T;V^*),L^2(0,T;V)} + \nu(\nabla y, \nabla v)_{L^2} = 0 \quad \forall\, v \in L^2(0,T;V),$$
$$(y(0) - y_0, v_0)_{L^2} = 0 \quad \forall\, v_0 \in H.$$

Lagrangian: $\lambda = (w, w_0) \in \Lambda = L^2(0,T;V) \times H$

$$L(y,u,\lambda) = J(y,u) + \langle y_t + y \cdot \nabla y - 1_{\Omega_c} u, w \rangle_{L^2(0,T;V^*),L^2(0,T;V)} + \nu(\nabla y, \nabla w)_{L^2}$$
$$+ (y(0) - y_0, w_0)_{L^2}.$$

Example: Optimal Control of 2D Navier Stokes Equations

TECHNISCHE UNIVERSITÄT DARMSTADT

Objective function: $J(y,u) := \frac{1}{2}\|y - y_Q\|^2_{L^2(Q)} + \frac{\alpha}{2}\|u\|^2_{L^2((0,T)\times\Omega_c)}$

State equation: $y \in Y = W(0,T), \quad C(y,u) = 0$ is given by

$$\langle y_t + y \cdot \nabla y - 1_{\Omega_c} u, v\rangle_{L^2(0,T;V^*),L^2(0,T;V)} + \nu(\nabla y, \nabla v)_{L^2} = 0 \quad \forall\, v \in L^2(0,T;V),$$
$$(y(0) - y_0, v_0)_{L^2} = 0 \quad \forall\, v_0 \in H.$$

Lagrangian: $\lambda = (w, w_0) \in \Lambda = L^2(0,T;V) \times H$

$$L(y,u,\lambda) = J(y,u) + \langle y_t + y \cdot \nabla y - 1_{\Omega_c} u, w\rangle_{L^2(0,T;V^*),L^2(0,T;V)} + \nu(\nabla y, \nabla w)_{L^2}$$
$$+ (y(0) - y_0, w_0)_{L^2}.$$

Adjoint equation: $L_y(y,u,\lambda) = 0$, $\lambda = (w, w_0) \in \Lambda$ is given by

$$(y - y_Q, v)_{L^2} + \langle v_t + v \cdot \nabla y + y \cdot \nabla v, w\rangle_{L^2(0,T;V^*),L^2(0,T;V)}$$
$$+\nu(\nabla v, \nabla w)_{L^2} + (v(0), w_0)_{L^2} = 0 \quad \forall\, v \in W(0,T).$$

Reduced derivative: $\widehat{J}'(u) = L_u(y,u,\lambda) = \alpha u - 1_{\Omega_c} w.$

Basic Trust Region Method 1

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min_{u \in U} \ \widehat{J}(u) := J(y(u), u) \quad \text{s. t.} \quad u \in U_{ad},$$

Trust region method:

$u_{k+1} = u_k + s_k$, where s_k is an approximate solution of the

Trust region problem at u_k:

$$\min_{s \in U} \ \widehat{q}_k(s) := \widehat{J}(u_k) + (\widehat{g}_k, s)_U + \frac{1}{2}\langle s, \widehat{H}_k s\rangle_{U,U^*}$$
$$\text{s. t.} \quad u_k + s \in U_{ad}, \quad \|s\|_U \le \delta_k$$

- $\widehat{g}_k$ reduced gradient: Riesz representation of $\widehat{J}'(u_k) = L_u(y_k, u_k, \lambda_k)$
- y_k solves state eq. $C(y_k, u_k) = 0$, λ_k solves adjoint eq. $L_y(y_k, u_k, \lambda_k) = 0$
- $\widehat{H}_k$ approximation of reduced Hessian $\widehat{J}''(u_k)$
- $\delta_k > 0$ trust region radius (adjusted by the trust region method)

Basic Trust Region Method 2

TECHNISCHE UNIVERSITÄT DARMSTADT

Choose $0 < \eta_1 < \eta_2 < 1$ and initial $u_0 \in U_{ad}$. Set $k := 0$.

1. Compute state $y_k = y(u_k)$, i.e. $C(y_k, u_k) = 0$.
2. Compute adjoint state λ_k, i.e. $L_y(y_k, u_k, \lambda_k) = 0$.
3. Compute reduced gradient $\widehat{g}_k$ as Riesz representation of $L_u(y_k, u_k, \lambda_k)$.
3. If $\|u_k - P_{U_{ad}}(u_k - \widehat{g}_k)\|_U \le \varepsilon_{tol}$: STOP
4. Compute s_k as inexact solution of the trust region problem

$$\boxed{\min \quad \widehat{q}_k(s) \quad \text{subject to} \quad \|s\|_U \le \delta_k, u_k + s \in U_{ad}.}$$

5. Compute new state $y_{k+1} = y(u_k + s_k)$.
6. **Accept or Reject Step:** If

$$\frac{J(y_k, u_k) - J(y_{k+1}, u_{k+1})}{\widehat{q}_k(0) - \widehat{q}_k(s_k)} \begin{cases} \ge \eta_1 & (u_{k+1}, y_{k+1}) := (u_k + s_k, y_{k+1}),\ \delta_{k+1} := \delta_k, \\ \ge \eta_2 & (u_{k+1}, y_{k+1}) := (u_k + s_k, y_{k+1}),\ \delta_{k+1} > \delta_k, \\ < \eta_1 & (u_{k+1}, y_{k+1}) := (u_k, y_k),\ \delta_{k+1} := \delta_k/2. \end{cases}$$

$k \leftarrow k + 1$. Goto 1.

Multilevel Trust Region Algorithm: Discretization

TECHNISCHE UNIVERSITÄT DARMSTADT

We assume for simplicity a conformal discretization:

$$Y_h \subset Y, \quad U_h \subset U, \quad \Lambda_h \subset \Lambda,$$

Let in iteration k a current grid and a control $u_k \in U_{h_k}$ be given.

Discretized state equation:

Compute $y_k = y_k^{h_k} \in Y_{h_k}$ by applying a convergent solver of $C(y, u_k) = 0$

$$C^{h_k} : Y_{h_k} \times U_{h_k} \to \Lambda_{h_k}^*, \qquad C^{h_k}(y_k, u_k) = 0.$$

Discretized adjoint equation:

Compute $\lambda_k = \lambda_k^{h_k} \in \Lambda_{h_k}$ by applying a convergent solver of $L_y(y_k, u_k, \lambda) = 0$.

Discrete reduced gradient:

$$\widehat{g}_k = \text{Riesz representation of } L_u(y_k, u_k, \lambda_k) \text{ in } U_h.$$

Criticality Measure

TECHNISCHE UNIVERSITÄT DARMSTADT

$U_{ad} \subset U$ closed, convex, $P_{U_{ad}}$ projection onto U_{ad}.

Criticality measure:

$$\|Pg_k\|_U := \|u_k - P_{U_{ad}}(u_k - \widehat{g}_k)\|_U$$

Discrete criticality measure:

$$\|P^h g_k\|_{U_h} := \|u_k - P^h_{U^h_{ad}}(u_k - \widehat{g}_k)\|_{U_h}$$

$P^h_{U^h_{ad}}$ projection onto U^h_{ad}.

Example: $U_{ad} = \{u \in U : a \le u \le b\}, \quad a, b \in L^\infty,\ a \le b.$

$$P_{U_{ad}}(u_k - \widehat{g}_k) = \max(a, \min(u_k - \widehat{g}_k, b))$$

Adaptive Multilevel Trust Region Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

Choose initial discretization (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$.

1. Compute discrete state y_k (sufficiently) exactly (if not already done).
2. Compute a discrete adjoint state λ_k, inexact reduced gradient $\widehat{g}_k$.

R1. Estimate $\|C(y_k, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k, u_k, \lambda_k)\|_{Y^*}$, $\|Pg_k - P^{h_k} g_k\|_U$, compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k\|_{U_{h_k}} \leq \varepsilon_{tol}$: STOP
4. Compute s_k as inexact solution of

$$\min \quad \widehat{q}_k^{h_k}(s) \quad \text{subject to} \quad \|s\|_{U_{h_k}} \leq \delta_k, u_k + s \in U_{ad}^{h_k}.$$

5. Compute discrete state $y_{k+1} = y^{h_k}(u_k + s_k)$ (sufficiently) exactly.
6. Accept or reject step by reduction ratio, choose δ_{k+1}. $k \leftarrow k + 1$. Goto 1.
If rejected, check inexact reduced gradient $\widehat{g}_k$, possibly refine and go to 1.

Outline

- Problem setting and basic concept
- Adaptive multilevel methods for time-dependent PDE-constrained optimization with control constraints
- Implementation of the error control criteria by error estimators
- Extension to reduced order models with error estimation
- Numerical results

Adaptive Multilevel Trust Region Algorithm

Choose initial discretization (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$.

1. Compute discrete state y_k (sufficiently) exactly (if not already done).
2. Compute a discrete adjoint state λ_k, inexact reduced gradient $\widehat{g}_k$.

R1. Estimate $\|C(y_k, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k, u_k, \lambda_k)\|_{Y^*}$, $\|P g_k - P^{h_k} g_k\|_U$, compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k\|_{U_{h_k}} \le \varepsilon_{tol}$: STOP
4. Compute s_k as inexact solution of

$$\boxed{\min \quad \widehat{q}_k^{h_k}(s) \quad \text{subject to} \quad \|s\|_{U_{h_k}} \le \delta_k, u_k + s \in U_{ad}^{h_k}.}$$

5. Compute discrete state $y_{k+1} = y^{h_k}(u_k + s_k)$ (sufficiently) exactly.
6. Accept or reject step by reduction ratio, choose δ_{k+1}. $k \leftarrow k + 1$. Goto 1.
If rejected, check inexact reduced gradient $\widehat{g}_k$, possibly refine and go to 1.

Adaptive Multilevel Trust Region Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

Choose initial discretization (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$.

1. Compute discrete state y_k (sufficiently) exactly (if not already done).
2. Compute a discrete adjoint state λ_k, inexact reduced gradient $\widehat{g}_k$.

R1. Estimate $\|C(y_k, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k, u_k, \lambda_k)\|_{Y^*}$, $\|P g_k - P^{h_k} g_k\|_U$, compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k\|_{U_{h_k}} \leq \varepsilon_{tol}$: STOP
4. Compute s_k as inexact solution of

$$\min \quad \widehat{q}_k^{h_k}(s) \quad \text{subject to} \quad \|s\|_{U_{h_k}} \leq \delta_k, u_k + s \in U_{ad}^{h_k}.$$

5. Compute discrete state $y_{k+1} = y^{h_k}(u_k + s_k)$ (sufficiently) exactly.
6. Accept or reject step by reduction ratio, choose δ_{k+1}. $k \leftarrow k + 1$. Goto 1.
   If rejected, check inexact reduced gradient $\widehat{g}_k$, possibly refine and go to 1.

Refinement Criteria

Step R1,R2: Estimate residuals in optimality conditions

Implement with fixed (unknown) constants $K_y, K_\lambda, K_u > 0$ the following criteria:

Check if

$$\begin{aligned}
\|C(y_k, u_k)\|_{\Lambda^*} &\leq K_y \|P^{h_k} g_k\|_{U_{h_k}} \\
\|L_y(y_k, u_k, \lambda_k)\|_{Y^*} &\leq K_\lambda \|P^{h_k} g_k\|_{U_{h_k}} \\
\|P g_k - P^{h_k} g_k\|_U &\leq K_u \|P^{h_k} g_k\|_{U_{h_k}}
\end{aligned}$$

Otherwise refine the grid for Y_{h_k}, Λ_{h_k}, U_{h_k} and goto Step 1.

Refinement Criteria

Step R1,R2: Estimate residuals in optimality conditions

Implement with fixed (unknown) constants $K_y, K_\lambda, K_u > 0$ the following criteria:

Check if

$$\begin{aligned} \|C(y_k, u_k)\|_{\Lambda^*} &\le K_y \|P^{h_k} g_k\|_{U_{h_k}} \\ \|L_y(y_k, u_k, \lambda_k)\|_{Y^*} &\le K_\lambda \|P^{h_k} g_k\|_{U_{h_k}} \\ \|P g_k - P^{h_k} g_k\|_U &\le K_u \|P^{h_k} g_k\|_{U_{h_k}} \end{aligned}$$

Otherwise refine the grid for Y_{h_k}, Λ_{h_k}, U_{h_k} and goto Step 1.

Assume that we have reliable error estimators $\eta_{y,h}$, $\eta_{\lambda,h}$, $\eta_{Pg,h}$ with

$$\begin{aligned} \|C(y_k, u_k)\|_{\Lambda^*} &\le C_1 \eta_{y,h_k} \\ \|L_y(y_k, u_k, \lambda_k)\|_{Y^*} &\le C_2 \eta_{\lambda,h_k} \\ \|P g_k - P^{h_k} g_k\|_U &\le C_3 \eta_{Pg,h_k} \end{aligned}$$

Refinement Criteria

TECHNISCHE UNIVERSITÄT DARMSTADT

Step R1,R2: Estimate residuals in optimality conditions

Implement with fixed constants $\tilde{K}_y, \tilde{K}_\lambda, \tilde{K}_u > 0$ the following criteria:

Check if

$$\begin{aligned} \eta_{y,h_k} &\leq \tilde{K}_y \| P^{h_k} g_k \|_{U_{h_k}} \\ \eta_{\lambda,h_k} &\leq \tilde{K}_\lambda \| P^{h_k} g_k \|_{U_{h_k}} \\ \eta_{Pg,h_k} &\leq \tilde{K}_u \| P^{h_k} g_k \|_{U_{h_k}} \end{aligned}$$

Otherwise refine the grid for Y_{h_k}, Λ_{h_k}, U_{h_k} and goto Step 1.

Assume that we have reliable error estimators $\eta_{y,h}$, $\eta_{\lambda,h}$, $\eta_{Pg,h}$ with

$$\begin{aligned} \| C(y_k, u_k) \|_{\Lambda^*} &\leq C_1 \eta_{y,h_k} \\ \| L_y(y_k, u_k, \lambda_k) \|_{Y^*} &\leq C_2 \eta_{\lambda,h_k} \\ \| P g_k - P^{h_k} g_k \|_U &\leq C_3 \eta_{Pg,h_k} \end{aligned}$$

Adaptive Multilevel Trust Region Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

Choose initial discretization (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$.

1. Compute discrete state y_k (sufficiently) exactly (if not already done).
2. Compute a discrete adjoint state λ_k, inexact reduced gradient $\widehat{g}_k$.

R1. Estimate $\|C(y_k, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k, u_k, \lambda_k)\|_{Y^*}$, $\|Pg_k - P^{h_k} g_k\|_U$, compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k\|_{U_{h_k}} \leq \varepsilon_{tol}$: STOP
4. Compute s_k as inexact solution of

$$\min \quad \widehat{q}_k^{h_k}(s) \quad \text{subject to} \quad \|s\|_{U_{h_k}} \leq \delta_k, u_k + s \in U_{ad}^{h_k}.$$

5. Compute discrete state $y_{k+1} = y^{h_k}(u_k + s_k)$ (sufficiently) exactly.
6. Accept or reject step by reduction ratio, choose δ_{k+1}. $k \leftarrow k + 1$. Goto 1.
If rejected, check inexact reduced gradient $\widehat{g}_k$, possibly refine and go to 1.

Adaptive Multilevel Trust Region Algorithm

Choose initial discretization (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$.

1. Compute discrete state y_k (sufficiently) exactly (if not already done).
2. Compute a discrete adjoint state λ_k, inexact reduced gradient $\widehat{g}_k$.

R1. Estimate $\|C(y_k, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k, u_k, \lambda_k)\|_{Y^*}$, $\|P g_k - P^{h_k} g_k\|_U$, compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k\|_{U_{h_k}} \leq \varepsilon_{tol}$: STOP
4. Compute s_k as inexact solution of

$$\min \quad \widehat{q}_k^{h_k}(s) \quad \text{subject to} \quad \|s\|_{U_{h_k}} \leq \delta_k, u_k + s \in U_{ad}^{h_k}.$$

5. Compute discrete state $y_{k+1} = y^{h_k}(u_k + s_k)$ (sufficiently) exactly.
6. Accept or reject step by reduction ratio, choose δ_{k+1}. $k \leftarrow k+1$. Goto 1.
If rejected, check inexact reduced gradient $\widehat{g}_k$, possibly refine and go to 1.

Trust Region Step

Step 4: Compute s_k as inexact solution of

$$\begin{aligned} &\min \quad \widehat{q}_k^{h_k}(s) = J(y_k, u_k) + (\widehat{g}_k^{h_k}, s)_{U_{h_k}} + \frac{1}{2}\left\langle s, \widehat{H}_k^{h_k} s \right\rangle_{U_{h_k}, U_{h_k}^*} \\ &\text{subject to} \quad \|s\|_U \leq \delta_k, u_k + s \in U_{ad}^{h_k}. \end{aligned}$$

that satisfies the **Generalized Cauchy decrease condition**

$$\widehat{q}_k^{h_k}(0) - \widehat{q}_k^{h_k}(s_k) \geq \kappa_q \|P^{h_k} g_k\|_{U_{h_k}} \min\{\|P^{h_k} g_k\|_{U_{h_k}}, \delta_k\}.$$

with $\kappa_q \in (0, 1)$

Possibilities: e.g.

- Projected negative gradient direction with Armijo linesearch (works always if Hessians are bounded, $\|H_k^{h_k}\| \leq R$, $R > 0$)
- Projected pcg-Newton step with Armijo linesearch

Adaptive Multilevel Trust Region Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

Choose initial discretization (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$.

1. Compute discrete state y_k (sufficiently) exactly (if not already done).
2. Compute a discrete adjoint state λ_k, inexact reduced gradient $\widehat{g}_k$.

R1. Estimate $\|C(y_k, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k, u_k, \lambda_k)\|_{Y^*}$, $\|Pg_k - P^{h_k} g_k\|_U$, compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k\|_{U_{h_k}} \leq \varepsilon_{tol}$: STOP
4. Compute s_k as inexact solution of

$$\min \quad \widehat{q}_k^{h_k}(s) \quad \text{subject to} \quad \|s\|_{U_{h_k}} \leq \delta_k, u_k + s \in U_{ad}^{h_k}.$$

5. Compute discrete state $y_{k+1} = y^{h_k}(u_k + s_k)$ (sufficiently) exactly.
6. Accept or reject step by reduction ratio, choose δ_{k+1}. $k \leftarrow k + 1$. Goto 1.
If rejected, check inexact reduced gradient $\widehat{g}_k$, possibly refine and go to 1.

Adaptive Multilevel Trust Region Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

Choose initial discretization (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$.

1. Compute discrete state y_k (sufficiently) exactly (if not already done).
2. Compute a discrete adjoint state λ_k, inexact reduced gradient $\widehat{g}_k$.

R1. Estimate $\|C(y_k, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k, u_k, \lambda_k)\|_{Y^*}$, $\|Pg_k - P^{h_k} g_k\|_U$, compare with $\|P^{h_k} g_k\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k\|_{U_{h_k}} \le \varepsilon_{tol}$: STOP
4. Compute s_k as inexact solution of

$$\min \quad \widehat{q}_k^{h_k}(s) \quad \text{subject to} \quad \|s\|_{U_{h_k}} \le \delta_k, u_k + s \in U_{ad}^{h_k}.$$

5. Compute discrete state $y_{k+1} = y^{h_k}(u_k + s_k)$ (sufficiently) exactly.
6. Accept or reject step by reduction ratio, choose δ_{k+1}. $k \leftarrow k + 1$. Goto 1.
   If rejected, check inexact reduced gradient $\widehat{g}_k$, possibly refine and go to 1.

Refinement Criterion for Inexact Reduced Gradient

TECHNISCHE UNIVERSITÄT DARMSTADT

In our "first optimize then discretize" approach we allow:

Independent adjoint solver for adjoint PDE

$\rightsquigarrow$ inexact reduced gradient $\widehat{g}_k$

Implement with arbitrary constant $\xi > 0$ the following criterion:

In Step 6: If step is not acceptable then check

$$|(\widehat{g}_k, s_k)_{U_{h_k}} - \langle \nabla \widehat{J}(u_k), s_k \rangle| \leq \xi \min\{\|P^{h_k} g_k\|_{U_{h_k}}, \delta_k\} \|s_k\|_{U_{h_k}}.$$

Otherwise refine the grids and go to step 1.

Criterion for Mesh Adaptation

Use a refinement strategy such that

$$\sum_{k=0}^{\infty}\left(J(y_{k+1}^{h_{k+1}}, u_{k+1}^{h_k}) - J(y_{k+1}^{h_k}, u_{k+1}^{h_k})\right) < \infty.$$

Meaning:

- Refine such that continuous problem is approached sufficiently fast
- Results of Nochetto, Siebert et al. indicate that standard refinement techniques satisfy the condition

Convergence Result

Theorem: (Ziems, S.U. SIOPT 2011, Ziems SIOPT 2013)

If standard assumptions hold uniformly on all generated grids then

- for $\varepsilon_{tol} > 0$ the adaptive multilevel trust region method with the given refinement criteria terminates finitely and
- for $\varepsilon_{tol} = 0$ the algorithm terminates finitely with a stationary point or produces a sequence with $u_k \in U_{ad}$, $k \in \mathbb{N}$, and

$$\liminf_{k\to\infty} \|C(y_k, u_k)\|_{\Lambda^*} + \|L_y(y_k, u_k, \lambda_k)\|_{Y^*} + \|P\widehat{g}_k\|_U = 0.$$

Local Refinement Strategy

TECHNISCHE UNIVERSITÄT DARMSTADT

Based on Elementwise Contributions to Error Estimators:

$$\eta_{C,h} = \left(\sum_{\text{Elements } T \text{ in triangulation}} \eta_{C,h,T}^2 \right)^{1/2}$$

$$\eta_{L_y,h} = \left(\sum_{\text{Elements } T \text{ in triangulation}} \eta_{L_y,h,T}^2 \right)^{1/2}$$

Typical Example for Refinement Strategy:

- **State:** Refine the $p\%$ elements with largest local errors $\eta_{C,h,T}$.
- **Adjoint:** Refine the $p\%$ elements with largest local errors $\eta_{L_y,h,T}$.

Suitable error estimators: e.g., Bank, Bartels, Becker, Carstensen, Funken, Hoppe, Rannacher, Siebert, Süli, Zienkiewicz, Zhu, ...

Outline

- Problem setting and basic concept
- Adaptive multilevel methods for time-dependent PDE-constrained optimization with control constraints
- Implementation of the error control criteria by error estimators
- Extension to reduced order models with error estimation
- Numerical results

Basic Idea for Including Reduced Order Models

- Approximate current adaptive FE-discretization by ROM.
- Compute optimization steps with the ROM by using the same adaptive optimization method as before (now: FE-model instead of PDE-model, ROM instead of FE-model).
- Estimate error between state/adjoint FE-model and state/adjoint ROM (e.g., Grepl, Patera 2005; Haasdonk, Ohlberger 2006; Tröltzsch, Volkwein 2009; Urban, Patera 2012).
- Instead of mesh refinement improve ROM.
- If ROM requires improvement, check if FE-model itself needs refinement.
- If necessary, refine FE-model adaptively as before and update the ROM.

For simplicity:

On the ROM level we use exact discrete adjoints, i.e. exact reduced gradients.

Example: Reduced Order Models by Proper Orthogonal Decomposition

TECHNISCHE UNIVERSITÄT DARMSTADT

- Let $C(y, u) = 0$ be given by

$$\langle y_t(t), v \rangle_{V^*,V} + a(y(t), v) = \langle Bu(t), v \rangle_{V^*,V} \quad \forall\, v \in V, t \in (0, T), \quad y(0) = y_0.$$

- **FE-Discretization (impl. Euler):** Find $y^{fe}(t_i) \in V_h$ with $y^{fe}(t_0) = y_{0,h}$,

$$\left\langle \frac{y^{fe}(t_i) - y^{fe}(t_{i-1})}{\Delta t_i}, v_h \right\rangle_{V_h^*,V_h} + a(y^{fe}(t_i), v_h) = \langle Bu(t_i), v_h \rangle_{V_h^*,V_h} \quad \forall\, v_h \in V_h.$$

Example: Reduced Order Models by Proper Orthogonal Decomposition

TECHNISCHE UNIVERSITÄT DARMSTADT

- Let $C(y,u) = 0$ be given by

$$\langle y_t(t), v\rangle_{V^*,V} + a(y(t),v) = \langle Bu(t), v\rangle_{V^*,V} \quad \forall\, v \in V, t \in (0,T), \quad y(0) = y_0.$$

- **FE-Discretization (impl. Euler):** Find $y^{fe}(t_i) \in V_h$ with $y^{fe}(t_0) = y_{0,h}$,

$$\left\langle \frac{y^{fe}(t_i) - y^{fe}(t_{i-1})}{\Delta t_i}, v_h \right\rangle_{V_h^*,V_h} + a(y^{fe}(t_i), v_h) = \langle Bu(t_i), v_h\rangle_{V_h^*,V_h} \quad \forall\, v_h \in V_h.$$

- **Construction of POD-Basis:**
  For the span of snapshots $S = \text{span}\{y^{fe}(t_1), \dots, y^{fe}(t_N)\}$ generate an ONB $\{\varphi_1, \dots, \varphi_m\}$ (POD-Basis) w.r.t. some $(\cdot,\cdot)_H$, such that $\text{span}\{\varphi_1, \dots, \varphi_l\}$, $l \le m$, is best approximation with dimension l of the snapshots.
- Choose l and set $V_h^r := \text{span}\{\varphi_1, \dots, \varphi_l\}$
- **POD-ROM (impl. Euler):** Find $y^r(t_i) \in V_h^r$ with $y^r(t_0) = y_{0,h}^r$,

$$\left\langle \frac{y^r(t_i) - y^r(t_{i-1})}{\Delta t_i}, \varphi_j \right\rangle_{V_h^*,V_h} + a(y^r(t_i), \varphi_j) = \langle Bu(t_i), \varphi_j\rangle_{V_h^*,V_h}, \quad j = 1, \dots, l.$$

Adaptive Multilevel Trust Region Algorithm with Reduced Order Models

TECHNISCHE UNIVERSITÄT DARMSTADT

Choose (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$. Setup initial state /adjoint ROMs.

1. Compute discrete FE state y_k^{fe} (sufficiently) exactly (if not already done).
2. Compute a discrete FE adjoint state λ_k^{fe}, inexact reduced gradient $\widehat{g}_k^{fe}$.

R1. Estimate $\|C(y_k^{fe}, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k^{fe}\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k^{fe}, u_k, \lambda_k^{fe})\|_{Y^*}$, $\|P g_k^{fe} - P^{h_k} g_k^{fe}\|_U$, compare with $\|P^{h_k} g_k^{fe}\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k^{fe}\|_{U_{h_k}} \leq \varepsilon_{tol}$: STOP

Adaptive Multilevel Trust Region Algorithm with Reduced Order Models

TECHNISCHE UNIVERSITÄT DARMSTADT

Choose (P^{h_0}), $u_0 \in U_{ad}^{h_0}$. Set $k := 0$. Setup initial state /adjoint ROMs.

1. Compute discrete FE state y_k^{fe} (sufficiently) exactly (if not already done).
2. Compute a discrete FE adjoint state λ_k^{fe}, inexact reduced gradient $\widehat{g}_k^{fe}$.

R1. Estimate $\|C(y_k^{fe}, u_k)\|_{\Lambda^*}$ and compare with $\|P^{h_k} g_k^{fe}\|_{U_{h_k}}$.
(If ratio too large, refine grid for Y_{h_k}, Λ_{h_k} adaptively. Goto 1.)

R2. Estimate $\|L_y(y_k^{fe}, u_k, \lambda_k^{fe})\|_{Y^*}$, $\|P g_k^{fe} - P^{h_k} g_k^{fe}\|_U$, compare with $\|P^{h_k} g_k^{fe}\|_{U_{h_k}}$.
(If ratio too large, refine grids adaptively. Goto 1.)

3. If $\|P^{h_k} g_k^{fe}\|_{U_{h_k}} \leq \varepsilon_{tol}$: STOP

ROM.1. Compute ROM state y_k^r exactly (if not already done).

ROM.2. Compute ROM adjoint state λ_k^r, reduced ROM gradient $\widehat{g}_k^r$.

ROM.R. If estimator$(\|C^{h_k}(y_k^r, u_k)\|_{\Lambda_{h_k}^*} \|L_y^{h_k}(y_k^r, u_k, \lambda_k^r)\|_{Y_{h_k}^*}) >$
$\min(K_r \|P^{h_k,r} g_k^r\|_{U_{h_k}}^2, \kappa_C \mathrm{Tol}^2)$ or if $\|P^{h_k,r} g_k^r\|_{U_{h_k}} \leq \varepsilon_{tol}^r$:

- Check and possibly improve FE-model by calling steps 1–3.
- Improve state/adjoint ROM. Set $\mathrm{Tol} := \mathrm{Tol}/2$. Goto ROM.1.

Adaptive Multilevel Trust Region Algorithm with Reduced Order Models

TECHNISCHE UNIVERSITÄT DARMSTADT

1.–3. ...

ROM.1. Compute ROM state y_k^r exactly (if not already done).

ROM.2. Compute ROM adjoint state λ_k^r, reduced ROM gradient $\widehat{g}_k^r$.

ROM.R. If estimator$(\|C^{h_k}(y_k^r, u_k)\|_{\Lambda_{h_k}^*} \|L_y^{h_k}(y_k^r, u_k, \lambda_k^r)\|_{Y_{h_k}^*}) >$ $\min(K_r \|P^{h_k,r} g_k^r\|_{U_{h_k}}^2, \kappa_C \mathrm{Tol}^2)$ or if $\|P^{h_k,r} g_k^r\|_{U_{h_k}} \leq \varepsilon_{tol}^r$:

- Check and possibly improve FE-model by calling steps 1–3.
- Improve state/adjoint ROM. Set Tol := Tol/2. Goto ROM.1.

ROM.3. Compute s_k as inexact solution of the ROM-based problem

$$\min \quad \widehat{q}_k^{h_k,r}(s) \quad \text{subject to} \quad \|s\|_{U_{h_k}} \leq \delta_k, u_k + s \in U_{ad}^{h_k}.$$

ROM.4. Compute ROM state $y_{k+1}^r = y^{h_k,r}(u_k + s_k)$ exactly.

ROM.5. Accept/reject step, choose δ_{k+1}. $k \leftarrow k + 1$. Goto ROM.1.

Observations and Features

TECHNISCHE UNIVERSITÄT DARMSTADT

- One can show that

$$|J(y^{h_k}(u),u) - J(y^{h_k,r}(u),u)| \le c\|C^{h_k}(y_k^r,u_k)\|_{\Lambda_{h_k}^*}\|L_y^{h_k}(y_k^r,u_k,\lambda_k^r)\|_{Y_{h_k}^*} + \text{H.O.T.}$$

$$J(y^{h_k,r}(u_k),u_k) - J(y^{h_k,r}(u_{opt}),u_{opt}) \ge c\|P^{h_k,r}g_k^r\|_{U_{h_k}}^2$$

  This motivates the criterion

$$\text{estimator}(\|C^{h_k}(y_k^r,u_k)\|_{\Lambda_{h_k}^*}\|L_y^{h_k}(y_k^r,u_k,\lambda_k^r)\|_{Y_{h_k}^*}) > \min(K_r\|P^{h_k,r}g_k^r\|_{U_{h_k}}^2, \kappa_C \text{Tol}^2)$$

- Convergence properties remain the same as for FE-based method.
- For nonlinear problems, application of Discrete Empirical Interpolation is helpful to obtain efficient POD-based ROMs (Barrault, Maday, Nguyen, Patera 2004, Chaturantabut, Sorensen 2010).
- Steps are computed based on the ROM, hence second order methods ($\widehat{H}_k = (\widehat{J}^r)''(u_k)$) are not expensive.

Related work: Pironneau, Polak 2000; Arian, Fahl, Sachs 2000, Volkwein 2011.

A Posterior Error Estimators for ROMs

TECHNISCHE UNIVERSITÄT DARMSTADT

Error Estimators:

- Residuals $\|C^{h^k}(y_k^r, u_k)\|_{\Lambda^*_{h_k}}$, $\|L_y^{h^k}(y_k^r, u_k, \lambda_k^r)\|_{Y^*_{h_k}}$ directly computable.
- Error is often devided in space- and time-errors (Debrabant, Lang 09).
- To avoid computations on the FE-grid, one can prepare the spatial residual computation "offline" to speed up later frequent evaluations (e.g., Grepl, Patera; Canuto, Tonn, Urban; Haasdonk, Ohlberger; Urban, Patera; Quarteroni, Rozza, Iapichino et al.).

A Posterior Error Estimators for ROMs

TECHNISCHE UNIVERSITÄT DARMSTADT

Error Estimators:

- Residuals $\|C^{h^k}(y_k^r, u_k)\|_{\Lambda_{h_k}^*}$, $\|L_y^{h^k}(y_k^r, u_k, \lambda_k^r)\|_{Y_{h_k}^*}$ directly computable.
- Error is often devided in space- and time-errors (Debrabant, Lang 09).
- To avoid computations on the FE-grid, one can prepare the spatial residual computation "offline" to speed up later frequent evaluations (e.g., Grepl, Patera; Canuto, Tonn, Urban; Haasdonk, Ohlberger; Urban, Patera; Quarteroni, Rozza, Iapichino et al.).

Model Improvement Strategies:

- Include snapshots of the state/adjoint for the current control.
- Maximize the ROM error estimator w.r.t. the control and include corresponding state snapshots Patera; deVore; Canuto, Tonn, Urban et al.; Quarteroni, Rozza, Iapichino et al.
- Further approaches e.g. by Arian, Fahl, Sachs 2006; Afanasiev, Hinze 2001; Volkwein 2011; Ghiglieri, S.U. 2012.

Extension to State Constraints

TECHNISCHE UNIVERSITÄT DARMSTADT

PDE-constrained Optimization Problem (PS):

$$\min_{y \in Y, u \in U} \; J(y,u) \quad \text{subject to} \quad C(y,u) = 0, \quad u \in U_{ad}, \quad y \in Y_{ad}.$$

$Y \subset \mathcal{C}^0(Q)$	State space,	$U_{ad} \subset U$	Convex, closed set,
U	Control space (Hilbert space),	$J : Y \times U \to \mathbb{R}$	Objective function,
		$C : Y \times U_{ad} \to \Lambda^*$	State equation.

$Y_{ad} = \{y \in \mathcal{C}^0(Q) \,:\, y_a \le y \le y_b\}$,
$y_a, y_b \in \mathcal{C}^0(Q)$, $y_a < y_b$.

Moreau-Yosida Regularization

TECHNISCHE UNIVERSITÄT DARMSTADT

Regularized Optimization Problem $(\mathrm{PS})_\gamma$:

$$\min_{y\in Y, u\in U} \quad J_\gamma(y,u) := J(y,u) + \gamma\left(\|(y_a - y)_+\|^3_{L^3(Q)} + \|(y - y_b)_+\|^3_{L^3(Q)}\right)$$

$$\text{subject to} \quad C(y,u) = 0, \quad u \in U_{ad}.$$

γ Regularization parameter

$(y_a - y)_+ = \max(0, y_a - y)$

Bergounioux, Hintermüller, Hinze, Kunisch, Meyer, Neitzel, Rösch, Tröltzsch, M. Ulbrich, Yousept,...

Extension of the Adapative Multilevel Method to State Contraints

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min_{y\in Y, u\in U} \quad J^\gamma(y,u) := J(y,u) + \gamma\left(\|(y_a - y)_+\|^3_{L^3(Q)} + \|(y-y_b)_+\|^3_{L^3(Q)}\right)$$
$$\text{subject to} \quad C(y,u) = 0, \quad u \in U_{ad}.$$

Notation: Let $a : \mathbb{R}_+ \to \mathbb{R}+$ be strictly monotone decreasing, $\lim_{\gamma\to\infty} a(\gamma) = 0$

- Apply the multilevel algorithm to $(\mathrm{PS})_{\gamma_k}$ while adjusting γ_k.
- Refine the grid if one of the following conditions is violated

$$\begin{aligned}
\eta_{y,h_k} &\le \tilde{K}_y \max(\|P^{h_k} g_k^{\gamma_k}\|_{U_{h_k}}, c_y a(\gamma_k)) \\
\eta_{\lambda,h_k} &\le \tilde{K}_\lambda \max(\|P^{h_k} g_k^{\gamma_k}\|_{U_{h_k}}, c_\lambda a(\gamma_k)) \\
\eta_{Pg,h_k} &\le K_u \max(\|P^{h_k} g_k^{\gamma_k}\|_{U_{h_k}}, c_u a(\gamma_k))
\end{aligned}$$

- Increase γ_k if $a(\gamma_k) > K_a \|P^{h_k} g_k^{\gamma_k}\|_{U_{h_k}}$

Convergence analysis: Bott, S.U., Ziems 2012

Outline

TECHNISCHE UNIVERSITÄT DARMSTADT

- Problem setting and basic concept
- Adaptive multilevel methods for time-dependent PDE-constrained optimization with control constraints
- Implementation of the error control criteria by error estimators
- Extension to reduced order models with error estimation
- Numerical results

Numerical Results for Parabolic Problem with Control Constraints

TECHNISCHE UNIVERSITÄT DARMSTADT

Test Problem: Boundary Control of semilinear parabolic PDE

$$\min J(y,u) := \frac{1}{2}\|y(T)-y_d\|^2_{L^2(\Omega)} + \frac{1}{2}\|y-y_Q\|^2_{L^2(Q)} + \frac{\alpha}{2}\|u\|^2_{L^2((0,T)\times\partial\Omega)}$$

$$\text{subject to} \quad \begin{aligned} & y_t - \Delta y + y^3 = 0 \quad \text{in } Q := (0,T)\times\Omega, \\ & \partial_n y = u - y^4 \quad \text{on } (0,T)\times\partial\Omega, \\ & y(0) = y_0 \quad \text{in } \Omega \\ & a \le u \le b \end{aligned}$$

$\Omega = [0,1]\times[0,1]$ unit square,
$T = 1$, $\alpha = 0.01$,
$y_0 \equiv 1$, $y_d \equiv 0.2$, $y_Q(t) \equiv 1 - 0.8t$,
$a \equiv 0$, $b \equiv 0.4$

Iteration History

TECHNISCHE UNIVERSITÄT DARMSTADT

Algorithm with ROM (POD-DEIM):

It.	Ref.	$\eta_{y,h}[t,s]$	$\eta_{\lambda,h}[t,s]$	$\|\widetilde{P}^h g_k\|_U$	(t,x)-grid	POD, DEIM
0		1.8e-4 , 4.6e-3	8.5e-3 , 5.7e-3	2.6e-1	36 × 289	14, 16
1		2.7e-4 , 9.4e-3	3.0e-2 , 2.8e-3	4.9e-2	36 × 289	14, 16
	λ t	2.9e-4 ,	1.0e-4 ,	4.9e-2	67 × 289	14, 17
5		9.7e-4 , 1.1e-2	7.2e-6 , 1.5e-4	1.0e-3	67 × 289	14, 17
	y s,t	2.1e-4 , 1.7e-3	1.1e-6 , 2.3e-5	1.1e-3	157 × 1089	24, 30
11	y s,t	3.3e-5 , 1.5e-3	9.6e-7 , 4.5e-5	1.5e-4	157 × 1089	24, 30
14		2.1e-5 , 1.6e-3	9.4e-7 , 4.7e-5	7.0e-5	157 × 1089	24, 30
	y s,t	1.3e-5 , 7.7e-4	6.2e-6 , 8.6e-6	4.3e-5	217 × 4225	
	y s,t	3.1e-6 , 1.9e-4	4.4e-6 , 2.0e-6	4.8e-5	438 × 16641	24, 33
19		1.2e-6 , 2.6e-4	8.4e-8 , 3.1e-5	2.1e-5	438 × 16641	24, 33
	ROM	1.3e-6 , 2.2e-4	8.4e-8 , 3.1e-5	1.3e-4	438 × 16641	49, 33
23		8.0e-7 , 2.2e-4	8.4e-8 , 3.1e-5	2.0e-5	438 × 16641	49, 33
	ROM	8.3e-7 , 1.7e-4	8.4e-8 , 3.1e-5	2.0e-5	438 × 16641	74, 33
24		8.4e-7 , 1.7e-4	8.4e-8 , 3.1e-5	7.9e-6	438 × 16641	74, 33
	ROM	8.9e-7 , 1.9e-4	1.0e-6 , 2.0e-6	7.9e-6	438 × 16641	122, 33
		9.6e-7 , 1.9e-4	1.8e-6 , 2.0e-6	7.9e-6	438 × 16641	

Algorithm without ROM:

similar grids: 36 , 67 , 156 , 216 , 441 points in time, same spatial grids.

23	STOP	9.5e-7 , 1.9e-4	1.7e-7 , 2.0e-6	8.2e-6	441 × 16641

Adaptive Multilevel Method with ROM: Efficiency

TECHNISCHE UNIVERSITÄT DARMSTADT

Adaptive POD-DEIM, only boundary nonlinearity y^4:

(time optimization)/(time state simulation on last grid) $= 3.7$

Adaptive POD-DEIM, boundary nonlinearity y^4 and distributed nonlinearity y^3:

(time optimization)/(time state simulation on last grid) $= 3.85$

We observe similar performance for flow control problems.

Application: Optimal Glass Cooling

- Cooperation with J. Lang, D. Clever and J.C. Ziems within SPP 1253
- Quality control of complex glass production processes
- Avoid cracks and other defects
- High temperature $\rightarrow$ radiation has to be taken into account
- Discretization of the state and adjoint equation with the space-time adaptive solver KARDOS (Lang, Clever)

Modelling of Radiative Heat Transfer

TECHNISCHE UNIVERSITÄT DARMSTADT

Temperature $T(x,t)$ and Intensity $I(x,t,\nu,s)$:

$$\begin{aligned} \rho_m\, c_m\, \partial_t T - \nabla \cdot (k_c \nabla T) &= -\int_{\nu_0}^{\infty} \int_{S^2} \kappa\,(\,B_m(T,\nu) - I\,)\, ds\, d\nu \\ \forall \nu > \nu_0 : \quad s \cdot \nabla I + (\sigma + \kappa) I &= \frac{\sigma}{4\pi} \int_{S^2} I\, ds + \kappa\, B_m(T,\nu) \end{aligned}$$

$$B_m(T,\nu) = \frac{n_m^2}{c_0^2} \frac{2\, h_P\, \nu^3}{e^{h_p \nu/(k_B T)} - 1} \quad \text{(Planck function)}$$

$x \in \Omega$ space, $t \in [0, t_e]$ time, $\nu \in (0, \infty)$ frequency, $s \in S^2$ direction
Control: ambient temperature $u(t)$.

Coupling with KARDOS

TECHNISCHE UNIVERSITÄT DARMSTADT

Gray Scale Glass cooling problem (Pinnau, Klar, Thömmes, et al.)

$$\min \frac{1}{2}\int_0^{t_e} \|T - T_d\|^2_{L^2(\Omega)} dt + \frac{\delta_e}{2}\|(T - T_d)(t_e)\|^2_{L^2(\Omega)} + \frac{\delta_u}{2}\int_0^{t_e} \|u - u_d\|^2_{L^2(\partial\Omega)} dt$$

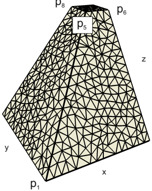
s.t.

$$\begin{aligned}
\partial_t T - k\Delta T - \frac{1}{3\kappa}\Delta\phi &= 0 && \text{in } \Omega \times (0, t_e] \\
-\frac{\epsilon^2}{3\kappa}\Delta\phi &= -\kappa\phi + 4\pi\kappa a T^4 && \text{in } \Omega \times (0, t_e] \\
kn \cdot \nabla T + \frac{1}{3\kappa} n \cdot \nabla\phi &= \frac{h}{\epsilon}(u - T) + \frac{1}{2\epsilon}4\pi a u^4 - \frac{1}{2\epsilon}\phi && \text{in } \partial\Omega \times (0, t_e] \\
\frac{\epsilon^2}{3\kappa} n \cdot \nabla\phi &= \frac{\epsilon}{2}(4\pi a u^4 - \phi) && \text{in } \partial\Omega \times (0, t_e] \\
T(0, x) &= T_0(x) && \text{in } \Omega
\end{aligned}$$

3D Optimal Control of Glass Cooling

TECHNISCHE UNIVERSITÄT DARMSTADT

Iteration history of the Second Order Multilevel Optimization Method



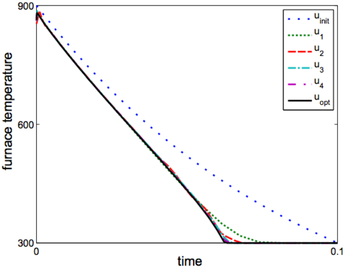
iteration	objective value	criticality measure	computing time
0	1.4602e+03	4.0245e+01	0.01%
1	5.4880e+02	2.8733e+00	0.04%
1	5.5664e+02	2.9758e+00	0.04%
2	5.4908e+02	3.4451e-01	0.35%
2	5.4507e+02	9.5386e-01	0.07%
3	5.4478e+02	3.0787e-01	1.05%
3	5.5319e+02	5.4405e-01	0.50%
4	5.5302e+02	1.0118e-01	9.96%
4	5.6995e+02	8.5377e-01	2.84%
5	5.6997e+02	6.8337e-02	85.10%

Bott, Clever, Lang, S.U., Ziems, Schröder 2014

3D Optimal Control of Glass Cooling

TECHNISCHE UNIVERSITÄT DARMSTADT

Control iterates u_k, $k = 0, \dots, 5$



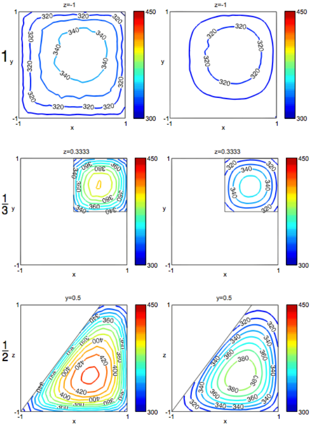
Initial space-time-grid: 25.000 DOF
Final space-time-grid: 1.755.000 DOF

Initial glass temperature Optimal glass temperature

$z = 1$

$z = \frac{1}{3}$

$y = \frac{1}{2}$



Thermistor Problem with State Constraints

TECHNISCHE UNIVERSITÄT DARMSTADT

Cf. Hömberg, Meyer, Rehberg, Ring; Meyer, Meinlschmidt

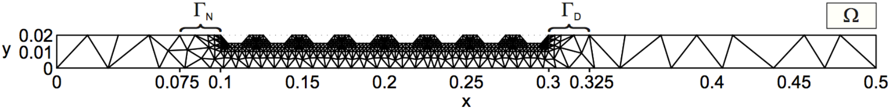
$$\min \frac{1}{2}\int_0^{t_e} \|T - T_d\|^2_{L^2(\Omega)}\,\mathrm{d}t + \frac{\delta_e}{2}\|(T - T_d)(t_e)\|^2_{L^2(\Omega)} + \frac{\delta_u}{2}\int_0^{t_e}(u - u_d)^2\mathrm{d}t$$

$$\begin{aligned}
\text{s.t.}\ \ C\rho\partial_t T - \nabla\cdot(\kappa\nabla T) &= (\sigma(T)\nabla\phi)\cdot\nabla\phi && \text{in } \Omega\times(0,t_e]\\
-\nabla\cdot(\sigma(T)\nabla\phi) &= 0 && \text{in } \Omega\times(0,t_e]\\
n\cdot(\kappa\nabla T) &= \alpha(T_l - T) && \text{in } \partial\Omega\times(0,t_e]\\
n\cdot(\sigma(T)\nabla\phi) &= C_u u && \text{in } \partial\Omega_N\times(0,t_e]\\
n\cdot(\sigma(T)\nabla\phi) &= 0 && \text{in } \partial\Omega_R\times(0,t_e]\\
\phi &= 0 && \text{in } \partial\Omega_D\times(0,t_e]\\
T(0,x) &= T_0 && \text{on } \Omega\\
T \le T_{max} &\ \text{a.e. in } \Omega\times(0,t_e]\\
0 \le u \le u_{max} &\ \text{a.e. in } (0,t_e]
\end{aligned}$$

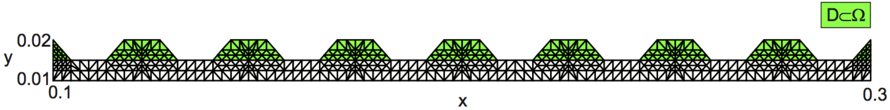
T temperature, ϕ electrical potential, u current

Thermistor Problem with State Constraints

Setup



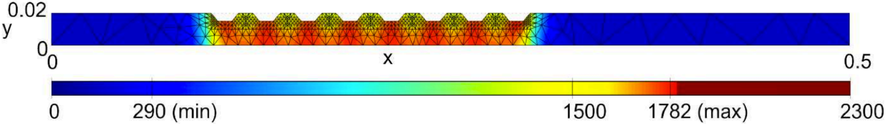
(a) Domain Ω with boundary Γ_N and Γ_D



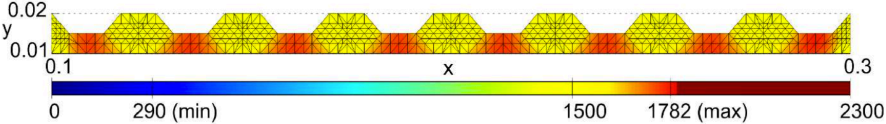
Bott, Clever, Lang, S.U., Ziems, Schröder 2014

Thermistor Problem with State Constraints

Optimal State (Temperature)



(a) computational domain Ω



(b) subdomain D

Bott, Clever, Lang, S.U., Ziems, Schröder 2014

Thermistor Problem with State Constraints

TECHNISCHE UNIVERSITÄT DARMSTADT

Iteration History

gamma	no. it.	objective	penalty term	max. pen.	crit. meas.	$\sum$dof	cpt
1.0e+0	0	3.2769e+5	3.2761e+5	9.2458e+2	1.4979e+2	2.37e+4	0.17%
1.0e+1	2	1.2535e+5	1.2534e+4	4.9127e+2	1.2846e+2	2.37e+4	2.66%
1.0e+2	3	1.3885e+4	1.3884e+2	2.4801e+2	3.4204e+1	2.64e+4	4.33%
1.0e+3	2	1.3390e+4	1.3389e+1	1.5877e+2	3.8208e+1	7.24e+4	8.85%
1.0e+4	3	2.5161e+3	2.5155e−1	6.3363e+1	1.5264e+1	6.97e+4	11.45%
1.0e+5	3	2.6217e+2	2.6115e−3	2.4467e+1	3.9273e+0	2.34e+5	44.57%
1.0e+6	2	2.5128e+2	2.5014e−4	1.7350e+1	3.3702e+0	2.32e+5	27.97%

Bott, Clever, Lang, S.U., Ziems, Schröder 2014

Conclusions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Multilevel optimization with adaptive discretization + ROMs + error estimation promising
- Retardation factor optimization/simulation ≤ 5 seems possible also for time-dependent problems
- Full second order methods applicable, since step is computed on ROM level

Topics for Future Work

- Integration of goal oriented instead of residual based error estimators
- Error estimation for ROMs (with Stefan Volkwein)
- Improvement strategies for ROMs
- ROMs for space-time adaptive discretizations
- Extension to robust optimization (under uncertainty)
- Applications, in particular flow and FSI problems, shape optimization